

DOCUMENT RESUME

ED 216 326

CS 006 664

AUTHOR Sartain, Harry W.
TITLE Frequency-Count of English Words Appearing in Books Read in Elementary Schools.
PUB DATE Aug 81
NOTE 27p.; Paper presented at the Annual Meeting of the European Conference on Reading (2nd, Joensuu, Finland, August 2-5, 1981).
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Basal Reading; *Child Language; Childrens Literature; Computational Linguistics; Elementary Education; *Reading Materials; *Reading Research; *Sight Vocabulary; *Word Frequency; *Word Lists; Word Recognition

ABSTRACT

To discover the truth about the extent of children's reading vocabularies, a project was undertaken at the Falk Laboratory School, University of Pittsburgh, to determine how many words first, second, and third grade children could recognize in print. A team of graduate students tabulated the words appearing in commonly used basal materials and in several library books that children read most often. This task resulted in a list of approximately 11,000 words. It was apparent, however, that a more complete list was needed for elementary school use. A new vocabulary word list, the Comprehensive Reading Vocabulary, itemized all of the words in the preprimers, primers, first grade readers, and second grade readers in nine commonly used reading series, as well as words from instructional and free-reading materials most frequently encountered by children in grades three, four, five, and six. When the computerized tabulation was completed, the total list consisted of 8,269,968 entries, of which 79,044 were different entries or "words." The first 10 words in the list in order of frequency include the 2 major articles, the most common conjunction in the English language, the 3 most common prepositions, 1 verb, and 3 pronouns. (The paper contains a list of the first 200 words classified according to their grammatical functions, as well as a list of the first 500 words classified according to their frequency.) (HOD)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

Harry W. Sartain

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

X This document has been reproduced as
received from the person or organization
originating it.

☐ Minor changes have been made to improve
reproduction quality.

• Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

FREQUENCY-COUNT OF ENGLISH WORDS
APPEARING IN BOOKS READ IN ELEMENTARY SCHOOLS

Harry W. Sartain

The size of the child's vocabulary always has been of great interest to parents and educators, because vocabulary is a good indicator of how successful the child is in learning. Its size tells whether or not the child has a good aptitude for using verbal symbol systems, and it helps to predict whether the child will continue to be successful in academic study.

Although any skillful observer can note whether two children are similar or somewhat different in their spoken vocabularies, the collection of precise data on oral vocabulary has been difficult because the measurement problems are so great. Consequently estimates of vocabulary size have differed considerably (Carroll, 1971). Probably the most accurate estimate of vocabulary size has been made by Professor Edgar Dale, who has spent fifty years in such study. He says that at the end of the first grade children will have an average vocabulary of 3,000 words and that they probably will learn about one thousand words a year after that (Dale, 1965, 948). He estimates, therefore, that the average high school senior will know between 14,000 and 15,000 words and that the college senior will know between 18,000 and 20,000. Some recent estimates are higher for the college senior.

Need for Reading Vocabulary Lists

It would seem logical that the vocabularies of various young children would be more similar than the vocabularies of college students and adults, because as people become more mature and choose different

vocations and professions, their vocabularies will become more specialized. It is equally logical to believe that the initial reading vocabularies to which different children are introduced can be fairly similar before they, themselves, have had many specialized experiences. But what should the words in a common reading vocabulary be?

One way to answer this question is to examine the vocabularies that have been used in books by the favorite authors of children and children's teachers. This has been done in the study of American English instructional and library materials to be described here. The presentation will include both a brief description of the process and the information about the vocabulary list that was developed. The "Comprehensive Reading Vocabulary" that resulted is still being refined and typed, but a list of the five hundred most frequent words is attached here.

Such a list should be of interest to at least four groups of people: (1) teachers of beginning reading in English-speaking countries, (2) teachers of English as a second language in other countries, (3) authors of books for children, and (4) publishers who supply instructional materials for schools. Insofar as other languages have vocabularies that parallel the English vocabulary, teachers of reading in other languages also might be interested in knowing which words appear most frequently in children's materials written in English.

Other Vocabulary Lists

This, of course, is far from the first elementary reading word list to be developed in English. Many lists have been compiled not only of reading vocabularies, but also of oral vocabularies and vocabularies in spelling and adult reading.

Oral vocabulary. The oral vocabularies of preschool children were studied as early as 1907 by Doran (1907, 401-38), but the first well-known vocabulary list in the United States was compiled by Horn in 1925. He tabulated 1,051 words commonly spoken by pre-first grade children and later added lists of words spoken by first grade children. The Horn list became a part of many other lists produced in future years. In 1928 the International Kindergarten Union published another list of 2,596 words spoken most frequently by kindergarten children. This, too, became a list that received major consideration by teachers and researchers.

In 1936 Buckingham and Dolch used a free association technique with 20,000 children from grades two through eight to compile another new vocabulary list. This was incorporated into earlier lists to form the Buckingham-Dolch "Combined Word List" of 9,520 words. Dale, who has focused attention on the comprehension of word meanings by children and adults, produced a list of 769 "easy" words in 1931, and in 1943 produced a second list of 2,926 "easy" words that were understood by eighty percent of fourth grade children (Dale, 1943). This widely used list has been revised and extended (Dale and O' Rourke, 1976), and is being revised again at this time.

Spelling vocabulary. Two or three hundred years ago the American colonists were teaching reading through the spelling method, a method

that they apparently had brought with them from Europe. This required that students learn to spell hundreds of syllables and thousands of words before they began textual reading. This method persisted in the United States in many schools throughout the nineteenth century. As the great inefficiency of the spelling method became apparent to teachers, researchers began to try to determine what words children really needed to learn to spell. In 1914 Jones compiled a list of 4,532 words found in children's themes, and in 1915 Ayres published data on the spelling difficulty of 1,000 common words in literary materials and adult correspondence. In 1921 Tidyman offered another list of the 3,000 words that occurred most frequently in children's compositions. In preparation for publishing a set of spelling textbooks, Horn, in 1926, offered another list and showed that children needed to learn to spell only 100 words to fulfill 58.5 per cent of their writing needs, and 3,000 words for 97.6 per cent of their writing.

The Ayers list and the Horn list were used extensively until the publication of a landmark study in 1945. During the great depression of the thirties the U.S. government provided employment for many people in cultural activities. One of these was a project directed by Rinsland to identify the words most frequently used in children's writing. Tabulating words from thousands of children's compositions and letters, the team sorted through six million running words to form a list of 25,362 different words. They found that 14,571 of these words were used more than thirty times, and Rinsland's A Basic Vocabulary of Elementary School Children was published in 1945. This list has had heavy use ever since.

Greene (1954) extended this work in a couple of extensive testing studies to determine at what grade levels children could spell correctly the long lists of these words that were used most frequently in writing. He published The New Iowa Spelling Scale (1954) based on his findings.

Adult reading vocabularies. Probably the Thorndike lists were the best-known lists ever produced. After examining both children's and adult literature, Thorndike published The Teacher's Word Book of Ten Thousand Words in 1921. In 1932 a revised list of 20,000 words was published. And in 1944, after adding a list of words from adult magazines, Thorndike and Lorge published The Teachers Word Book of 30,000 words. This list has been used in compiling most of the shorter lists since that time.

More recently Kucera and Francis (1967) used computer technology to produce their Computational Analysis of Present-Day American English. It was based on a variety of adult reading materials and provided the frequencies for 50,406 words.

Children's reading vocabularies. Sholty did one of the earliest studies of American children's reading vocabularies in 1912, and Gates in 1925 drew upon information from several other studies to produce the 1500-word Gates Reading Vocabulary for Primary Grades. Dolch, in turn, used the Gates list and several others to formulate his list of 220 service words in 1936. He indicated that these 220 words made up approximately two-thirds of the running words that occurred in the reading materials of the first three grades. His list of 220 service words still is used regularly by remedial reading teachers and primary teachers.

Stone analyzed primary reading materials and published a Graded Vocabulary for Primary Reading in 1941, while Krantz published The Author's Word List for Primary Grades in 1945. In 1952 Bryan concluded that most word lists were inadequate, because his study of 2,500 elementary school children's vocabularies revealed that sixth grade children generally knew 10,000 words, many of which were not in the published reading lists.

Shortly thereafter, Kyte (1953) combined words from common lists in preparing A Core Vocabulary in the Language Arts, and Penn (1953) selected 8,802 words from fourteen different lists to provide a graded word list for grades one through six. She found that the words that occurred most frequently in reading materials were taught most often in the lower grades, with less frequent words taught at higher levels. She commented on the lack of specific standards for judging the grade placement of words (Penn 1953). In 1961 Fullmer and Kolson, however, selected from the various lists 184 words that they recommended for use in initial reading instruction.

The most complete reading vocabulary list prepared to date was published by Carroll, Davis, and Richman in 1971. It was based on 10,000 500-word samples drawn from 1,045 texts to which students were exposed in grades three through nine. The textual materials included textbooks, workbooks, novels, poetry, nonfiction, encyclopedias, and magazines. From a total of 5,088,721 running words, the authors listed 86,741 different words both alphabetically and by rank.

Another recent computerized list was published by Harris and Jacobson in 1972. They tabulated the words from one hundred twenty-seven recent elementary textbooks, including six reading series and two series in each of the common content fields. They divided the total list of 7,613 different words into a general vocabulary list, graded lists for each level, and lists for each field of elementary study. Both the lists by Harris and Jacobson and by the Carroll group have great value for teachers and writers. The Harris list provides the most information for teaching of beginning readers, while the Carroll list is most useful to teachers at intermediate levels. The great size and ^{the} detail of/latter list makes it somewhat cumbersome for everyday use, however.

Development of a New List

Initial work on the new Comprehensive Reading Vocabulary by Sartain and Feldman was already well advanced when the lists by Harris and by Carroll were published. At that time a decision was made that this new list contained enough different material so that its development should continue. Specifically, the list contains words from more library materials than does the Harris list and it contains a full sampling of primary materials not included in the Carroll list. Therefore it can either support the findings presented from the other two vocabulary studies or it can extend them.

This list, too, was computerized. Changes in the computer system during the development of the list, however, interfered greatly with the project. Also the key-punching of the great amount of reading material that was included was very time-consuming and had to be drawn out

because of budget limitations. Consequently it has taken more time to produce this list than some others of a similar nature.

Background for the list. During the early sixties the book Why Johnny Can't Read, by Rudolph Flesch (1955) resulted in much critical attention to reading instruction in the United States. This was followed by a publication on What Ivan Knows that Johnny Doesn't (Trace, 1961) in which the author conveyed the opinion that American children can read only those words taught in basal readers. This revealed a great ignorance of the amount of reading that children are expected to do in other textbooks, in library books, and in various periodicals. The American public, however, is fairly gullible concerning criticisms of education, so these unfounded statements were quoted far and wide.

In an effort to discover the truth about the extent of children's reading vocabularies a project was undertaken at the Falk Laboratory School, University of Pittsburgh, to determine how many words first, second, and third grade children could recognize in print. First, of course, it was necessary to have a word list that would represent all of the words that such children might have been exposed to and have had an opportunity to learn. No such list was available in published form, since the only extensive list was the too-complete Lorge-Thorndike list of 30,000 words in adult reading. Therefore a team of graduate students set out to tabulate the words appearing in commonly used basal materials and in several library books that children read most often. This tedious task resulted in a list of approximately 11,000 words.

The words were listed in the order in which they appeared by grade level so that the list could be used as a sight vocabulary test

for children at the three grade levels. At the first grade level it was found that the children's reading vocabularies ranged from three words for the most immature pupil to more than 5,000 words for the most advanced. Many children could read more than three or four thousand words by the end of the first grade. By the end of the third grade a number of the children could read as many as nine thousand words (Sartain and Slick 1968). This certainly refuted the claim that American children could read only the fairly limited numbers of words presented in basal readers. It was apparent, however, that if a more adequate word list were available, it should include additional words that some of the children could read. This need for a more complete list for elementary school use led to the vocabulary study reported here.

Materials included. At the beginning of the new word list development Bird (1972) and Heimberger (1972) did a survey of a stratified sample of school districts throughout the United States to determine what textbooks in reading, language, mathematics, social studies, and science were most frequently used in the elementary schools. Then they itemized on computer keypunch cards all of the words in the preprimers, primers, first grade readers, and second grade readers in nine commonly used reading series that had been published in the late sixties. (These included two of the so-called "linguistic" series for beginning reading but not continued in higher grades.) Also they entered all the words from the first and second level books in other commonly used textbooks as follows: two language series, two mathematics series, three science series, and one social studies series. (The differences in numbers of series reflected both their relative variety of content and the numbers

of textbooks to which schools said primary children were exposed. For example, children were more likely to be exposed to more than one science textbook than to more than one social studies textbook.) Bird and Heimberger also used the findings from a library survey done in the research office at the Falk Laboratory School to select twenty-nine library books frequently read in the first two grades; the vocabularies of these books were entered into the computer system.

From this total of ninety books and booklets, Bird and Heimberger tabulated 444,794 running entries, of which 7,411 were different words.

With the assistance of others in the Falk School Research Office, Sartain and Feldman continued to feed into the computer the instructional and free-reading materials most frequently encountered by children in Grades three, four, five, and six. This increased the reading material input to include seven reading series, two language series, two mathematics series, three science series, two social studies series, 136 library fiction books, 68 nonfiction library books, 58 copies of frequently-read children's periodicals, and 258 issues of children's newspapers/magazines. Except for the reading series, the numbers of textbooks, library books, periodicals, and newspapers were in proportion to the numbers that teachers who were surveyed said that their children usually read.

The Sartain-Feldman List

When the computerized tabulation was completed, the total list consisted of 8,269,968 entries, of which 79,044 were different entries, or different "words." Because the keypunchers typed in everything that appeared on the printed pages, these entries included the separate letters of the alphabet used in outlines and other non-word letter combinations

used in various types of explanatory writing. Also, in spite of careful supervision, the keypunchers made a few typographical errors. Since the exact number of such non-word entries has not yet been counted, it seems appropriate to round-off the total list to approximately 8,200,000 running words and 79,000 different words.

The first ten words. The first ten words in the list in order of frequency are: the, and, a, of, to, in, you, is, he, and it. Not surprisingly, these include the two major articles, the most common conjunction in the English Language, the three most common prepositions, one verb, and three pronouns, but no nouns. Except for the articles, it is quite possible that the equivalent words in other languages also would occur with equally high frequency. The numbers of articles would differ greatly, however, because of the use of masculine, feminine, singular, and plural articles in some European languages. These numbers would also differ in Asiatic and African languages, which frequently use no articles.

The first two hundred words. In order to consider the types of words that young children must learn to read because of their high frequency, the first two hundred words have been classified in Table 1 according to their grammatical functions.

Table 1

TWO HUNDRED MOST FREQUENT WORDS CLASSIFIED BY GRAMMATICAL FUNCTION

<u>Art.</u>	<u>Noun</u>	<u>Pro- noun</u>	<u>Verb</u>	<u>Adjective</u>	<u>Adverb</u>	<u>Quali- fier</u>	<u>Prep.</u>	<u>Conj.</u>
<u>First Fifty Words:</u>								
a		he	are	all	how		about*	and
the		her	be	one	not		at	as
		his	can		then		by	but
		I	do		there		for	or
		it	had		when		from	so*
		she	have				in	
		that	is				on	
		they	said				of	
		this	was				out	
		we	were				to	
		what*	will				up	
		you					with	
		your						
<u>Second Fifty Words:</u>								
an	back*	him	been	each*	if	more	down	
	day	me	come	first	no	than	into	
	people	my	could	good	now	very	over	
	time	their	did	little		just*		
	water	them	find	many				
	way	these	get	new				
		which	go	only				
		who	has	other				
			know	some				
			like*	two				
			look					
			made					
			make					
			see					
			would					
<u>Third Fifty Words:</u>								
	home	its	asked	any	again	most	after	because
	house	our	called	big	away	much	around	
	man		came	great	here	too	before*	
	mother		does	long	where*		off	
	number		don't	Mr.	why		through	
	place		help*	old				
	things		looked	right				
	word		may	well*				
	words		must					
			put					
			take					
			tell					
			think					
			use					
			went					
			work					
			write					

Table 1 (continued)

TWO HUNDRED MOST FREQUENT WORDS CLASSIFIED BY GRAMMATICAL FUNCTION

<u>Art.</u>	<u>Noun</u>	<u>Pro- noun</u>	<u>Verb</u>	<u>Adjective</u>	<u>Adverb</u>	<u>Quali- fier</u>	<u>Prep.</u>	<u>Conj.</u>
<u>Fourth Fifty Words:</u>								
	air	it's*	began	another	also		below	
	boy	us	found	different	never			
	children		give	even*				
	city		going	every				
	father		got	last				
	head		let	left				
	land		own*	Mrs.				
	men		read	next				
	name		saw*	same				
	paper		say	small				
	part		set*	still				
	school		should	three*				
	sentence		thought*					
	something		took					
	story		used					
	world		want					
	years							

*These words could have been listed as other parts of speech with equal appropriateness.

The articles a and the occur among the first fifty words, and an is the first word in the second fifty. Obviously they must be taught very early.

The pattern of occurrence of nouns is interesting: none among the first fifty words, six among the second fifty, nine among the third fifty, and seventeen among the fourth fifty. This suggests that beginning readers may need to learn very few nouns at first, but that they will need to learn a large number soon thereafter.

The pattern for pronoun occurrence is exactly opposite from that for nouns: thirteen among the first fifty words, eight among the second fifty, two among the third, and two among the fourth fifty words. Obviously beginning readers must learn to read the pronouns in a hurry.

Verbs occur much more regularly among the first two hundred words than do nouns and pronouns: eleven among the first fifty, fifteen among the second, seventeen among the third, and sixteen among the fourth fifty. The reason for this is obvious: sentences, even of the simplest type, cannot be constructed without verbs. These first verbs include, incidentally, many of the forms of to be and to have that will be read as auxiliary verbs in the future.

The modifiers--adjectives, adverbs, and qualifiers--tend to increase gradually in use in each group of fifty words. There are two, ten, eight, and twelve adjectives; five, three, five, and two adverbs; and zero, four, three, and zero qualifiers among the groups of fifty words in the first two hundred. (The qualifiers, incidentally, are those words, formerly included among adverbs, that modify adjectives and other adverbs.) Apparently children will need to learn to read quite a few adjectives and a few adverbs right from the beginning and continuously thereafter for awhile.

The pattern of occurrence of the connective words--prepositions and conjunctions--is much like that of the pronouns. There are twelve prepositions among the first fifty words, with three, five, and one thereafter. There are five conjunctions among the first fifty, followed by zero, one, and zero. Obviously beginning readers need to cope with most of the prepositions and the simplest conjunctions in English from the beginning of their study of reading.

Looking at the proportions of words serving different grammatical functions in elementary reading, one can see that among the most frequently occurring words there is a heavy proportion of verbs, an increasing number of nouns, adjectives, and adverbs, and most of the common pronouns,

prepositions, and conjunctions.

The first five hundred words. A cursory scanning of the five hundred words used most frequently in elementary English reading shows that the load of nouns and verbs continues to become heavier and heavier. The number of adjectives and adverbs continues to increase also, while the number of additional connective words is quite small.

Another way to look at the list is to consider the relative frequencies of individual words and groups of words as proportions of the total list. As other lists have shown, some words are repeated so often that they constitute a large proportion of all words that children read. For example, a simple calculation reveals that, based on its frequency of occurrence in the total running vocabulary in elementary materials, the word the has a learning utility of 6.4%. That is, its frequency constitutes 6.4% of the total running vocabulary, so learning the word would be very helpful to the child in beginning reading. The words and, a, of and to have utility values of 2.7%, 2.6%, 2.3%, and 2.3% respectively. Because of the rapid decrease in frequency of other words in the list, however, most would have utility values of only a small fraction of one per cent.

The child who learns the first five hundred words in the "Comprehensive Reading Vocabulary" will know more than two-thirds of the running vocabulary in elementary materials. Table 2 shows the percentages of the total running vocabulary that are added as the child masters each group of high-frequency words.

Table 2

CUMULATIVE WORD TOTALS AND PERCENTAGES
IN THE COMPREHENSIVE READING VOCABULARY

<u>Word Frequency Group Included</u>	<u>Percentage Added</u>	<u>Cumulative Word Total</u>	<u>Cumulative Percent of Total Vocabulary</u>
Ten Most Frequent	22.2%	10 words	22.2%
Next Forty	17.8	50 words	40.0
Second Fifty	8.8	100 words	48.8
Next Ten		110 words	50.0
Third Fifty	5.1	150 words	53.9
Fourth Fifty	3.5	200 words	57.4
Fifth Fifty	2.6	250 words	60.0
Sixth Fifty	2.1	300 words	62.1
Fourth Hundred	3.2	400 words	65.3
Fifth Hundred	2.5	500 words	67.8
Next 78,500 (approx.)	32.2	79,000 (approx.)	100.0

This table indicates that a child needs to learn only ten words in order to be able to read twenty-two percent of all the running words encountered in the most common elementary school textbooks, library books, and periodicals. While he must learn less than fifty words in order to read the thirty-two percent of vocabulary most often repeated in elementary materials, he must learn seventy-eight thousand words in order to read the thirty-two per cent of the vocabulary that appears least frequently in elementary reading materials! This proves beyond doubt the need for educators to know the most frequent words in printed materials for children so that those words can be taught early in each child's school experience.

How important are the different groups of words in teaching beginning reading or remedial reading? Considering how frequently they are needed by elementary readers of the English Language, it appears that teachers

should consider the first fifty words to be a "Maximum Priority List," the next two hundred and fifty to be a "Major Priority List," and the fourth and fifth hundreds to be a "High Priority List." Children and students of English as a second language who have learned to read these five hundred words will know more than two-thirds of the vocabulary needed to read materials at the sixth grade level.

They also will have made a good beginning in learning the reading vocabulary they need for more advanced communication. It has been estimated that the average American high school graduate recognizes approximately 15,000 words when he is listening or reading and that the average college graduate recognizes approximately 30,000 words (Feinsilber and Mead 1980). Of course, students who are above average will know a considerably larger number of the more than half a million words in the English Language. Obviously when beginning readers have learned all of the 79,000 words in the "Comprehensive Reading Vocabulary," they will have mastered enough to show that they are progressing well on the road toward adult literacy.

The First Five Hundred Words
in
A COMPREHENSIVE READING VOCABULARY

18

By Harry W. Sartain and Renee Feldman
(with David Bird, Linda Benedict, Mary Heimberger, Betty Horodezky)

THE FIRST TEN (with frequencies) 22.2% of all running words

a	212,194	is	85,871	the	532,369
and	221,272	it	83,515	to	192,555
he	84,442	of	193,377	you	91,498
in	141,574				

THE NEXT FORTY (with frequencies) 17.8% of all running words

about	22,934	her	30,965	then	21,340
all	31,844	his	47,625	there	26,161
are	45,501	how	24,664	they	45,421
as	46,825	I	73,246	this	34,686
at	41,678	not	31,487	up	25,320
be	35,168	on	61,720	was	72,770
but	33,881	one	33,410	we	25,894
by	26,696	or	26,958	were	27,230
can	26,901	out	25,350	what	37,512
do	26,270	said	43,736	when	26,304
for	63,749	she	38,533	will	21,286
from	34,101	so	22,745	with	45,158
had	37,040	that	71,918	your	24,112
have	35,642				

THE SECOND FIFTY (with frequencies) 8.8% of all running words

an	20,973	into	18,492	over	12,391
back	12,169	just	12,627	people	13,976
been	11,007	know	12,370	see	17,399
come	10,263	like	18,890	some	18,348
could	16,741	little	15,069	than	10,919
day	10,628	look	10,342	their	17,289
did	15,699	made	11,245	them	19,440
down	15,203	make	13,912	these	13,440
each	18,181	many	17,666	time	13,804
find	12,081	me	16,496	two	14,003
first	11,502	more	14,596	very	10,626
get	13,855	my	20,823	water	12,530
go	13,303	new	12,698	way	11,415
good	10,648	no	16,671	which	14,420
has	13,452	now	14,787	who	12,152
him	20,529	only	10,278	would	19,203
if	20,064	other	15,334		

Fifty per cent of the total running words that appear in the most commonly used elementary textbooks, library books and magazines are included when the following ten words are added to the first one hundred: came, long, where, think, went, words, after, use, man, why.

THE THIRD FIFTY (Frequencies: 6945-10,214) 5.1% of running words

after	came	looked	old	too
again	does	man	our	use
any	don't	may	place	well
around	great	most	put	went
asked	help	mother	right	where
away	here	Mr.	take	why
because	home	much	tell	word
before	house	must	things	words
big	its	number	think	work
called	long	off	through	write

THE FOURTH FIFTY (Frequencies: 4838-6923) 3.5% of running words

air	every	left	read	still
also	father	let	same	story
another	found	men	saw	thought
began	give	Mrs.	say	three
below	going	name	school	took
boy	got	never	sentence	us
children	head	next	set	used
city	it's	own	should	want
different	land	paper	small	world
even	last	part	something	years

THE FIFTH FIFTY (Frequencies: 3878-4824) 2.6% of running words

along	enough	kind	once	soon
always	ever	knew	page	such
animals	eyes	life	picture	thing
book	feet	light	ran	those
boys	few	live	red	told
didn't	five	might	river	tree
dog	food	miss	room	under
door	hand	need	sea	while
earth	I'm	night	sentences	white
end	keep	oh	show	yes

THE SIXTH FIFTY (Frequencies: 3203-3875) 2.1% of running words

across	can't	heard	play	today
almost	country	high	run	together
am	days	I'll	sat	true
answer	eat	important	seen	try
being	face	large	side	until
best	far	line	sometimes	using
better	fish	looking	sound	wanted
between	following	miles	states	without
both	four	morning	sun	year
box	hard	plants	sure	young

THE FOURTH HUNDRED (Frequencies: 2299-3192)

3.2% of running words

above	done	himself	nothing	started
against	family	hot	numbers	stood
ago	fast	idea	ocean	stop
America	feel	Indians	often	stopped
anything	felt	inside	ones	table
ball	fire	John	open	talk
bed	form	king	order	ten
behind	friends	learn	others	tens
black	front	letter	parts	that's
blue	fun	living	pictures	times
books	gave	makes	point	top
call	girl	making	really	toward
car	girls	Mary	rock	trees
cat	gone	mean	second	turn
change	green	means	seemed	turned
cold	ground	money	several	united
course	grow	moon	since	voice
cried	half	move	south	ways
cut	hands	near	space	whole
dark	hear	north	special	wind

THE FIFTH HUNDRED (Frequencies: 1846-2296)

2.5% of running words

American	dry	Indian	questions	soil
animal	during	Jack	quite	stand
ask	early	kept	rain	start
bear	everyone	kinds	reading	state
beautiful	everything	later	ready	stay
became	example	learned	remember	street
bird	fine	less	rest	study
birds	free	let's	road	suddenly
body	friend	letters	round	summer
bring	full	lived	says	through
brought	game	map	scientists	Tom
brown	happened	mark	send	town
built	happy	maybe	shall	tried
cities	held	mind	ship	uncle
class	Henry	needed	short	walked
comes	hold	numeral	shown	warm
coming	horse	piece	shows	watch
copy	hundred	places	six	west
couldn't	ice	plant	slowly	winter
doing	inches	please	snow	yet

These words are from a COMPREHENSIVE READING VOCABULARY of approximately 79,000 words drawn from a corpus of 8,269,968 entries tabulated by computer from 105 reading textbooks and booklets, 54 textbooks in other fields, 136 fiction books, 68 nonfiction library books, 58 copies of children's periodicals, and 258 issues of news magazines frequently read by children in Grades 1-6 in the early seventies. A large proportion of the funding for the project came from the Falk School Research Office, University of Pittsburgh.

BIBLIOGRAPHY AND REFERENCES

- Ayers, Leonard P. Measurement of spelling ability. New York: Russell Sage Foundation, 1915.
- Beck, Mae M. The vocabulary of first readers, Peabody Journal of Education, Jan. 1929, VI, 224-7.
- Betts, Emmett A. A study of the vocabularies of first grade basal readers, Elementary English Review, Oct. 1938, XVI, 65-9.
- Bird, David G. An analysis of the grapheme-phoneme patterns in the vocabulary of library books and textbooks frequently read by early primary children. Unpublished doctoral dissertation, University of Pittsburgh, 1972.
- Brittain, Francis J., and Fitzgerald, James A. The vocabulary and spelling errors of second grade children's themes, Elementary English Review, Feb. 1942, XIX, 43-50.
- Bryan, Fred E. The size of children's vocabularies. Unpublished Ed.D. dissertation, School of Education, University of Pittsburgh, 1952.
- Buckingham, Burdette R., and Dolch, Edward W. A combined word list. New York: Ginn and Company, 1936.
- Carroll, John B. Development of native language skills beyond the early years. In Carroll E. Reed. The learning of language. New York: Appleton-Century-Crofts, 1971, 97-156.
- Carroll, John B.; Davis, Peter; and Richman, Barry. The American heritage word frequency book. Boston: Houghton Mifflin Co., 1971.
- Carson, Louise G. The use of preprimers: a teacher's point of view, Elementary School Journal, Dec. 1943, XLIII, 225-33.
- Cattell, J. McK. Time and space in vision, Psychological Review, 1900, VII, 325-343 (Probably reference, supplied by Huey, 1908, 434.)
- Child Study Committee of the International Kindergarten Union. A study of the vocabulary of children before entering first grade. Washington, D.C.: International Kindergarten Union, 1928.
- Cianflone, Ralph. Comparative difficulty of beginning reading vocabulary: set V (nouns). Unpublished doctoral dissertation, University of Pittsburgh, 1976.
- Cogar, Roy. Comparative difficulty of beginning reading vocabulary: set III (word length). Unpublished doctoral dissertation, University of Pittsburgh, 1974.

- Craig, James C. Readability of children's literature, Report of the Tenth Annual Conference on Reading, University of Pittsburgh, 1954, 113-124.
- Crouse, Russell. Comparative difficulty of beginning reading vocabulary: set IV (verbs). Unpublished doctoral dissertation, University of Pittsburgh, 1976.
- Dale, Edgar. Easy word list. Columbus, Ohio: The Ohio State University, 1943.
- Dale, Edgar, and Eichholz, Gerhard. Children's knowledge of words: an interim report. Columbus, Ohio: The Ohio State University, 1960.
- Dale, Edgar, and O'Rourke, Joseph. The living word vocabulary. Elgin Illinois: Dome Incorporated, 1976.
- Davis, Frederick B. Psychometric research on comprehension in reading, Reading Research Quarterly, Summer 1972, 628-78.
- Daw, Seward E. A controlled vocabulary for the middle grades. Unpublished Ph.D. dissertation, School of Education, University of Pittsburgh, 1940.
- Dolch, Edward W. A basic sight vocabulary, Elementary School Journal, Feb. 1936, XXXVI, 456-60.
- Dolch, Edward W. Needed vocabulary, Elementary English, Dec. 1960, XXXVII, 530-4.
- Door, John H. A study of word recognition in the first grade of the Monongahela public schools. Unpublished master's degree thesis, University of Pittsburgh, 1933.
- Durr, William K. A computer study of high frequency words in popular trade juveniles. East Lansing, Michigan. Michigan State University, 1970.
- Erdmann, B., and Dodge, R. Psychologische untersuchungen uber das lesen, auf experimenteller grundlage. Halle, 1898. (From Huey, 1908, 434.)
- Feinsilber, Mike, and Mead, William B. American averages. New York: Doubleday and Co., 1980. Quoted by the authors in Amazing American averages, Reader's Digest, July 1981, 115.
- Figurel, J. Allen. The vocabulary of underprivileged children. Unpublished Ph.D. dissertation, School of Education, University of Pittsburgh, 1948.
- Fitzgerald, James A. An integrating basic communication vocabulary, Elementary English, Mar. 1963, XL, 283-9.

- Fitzgerald, James A. The vocabulary and spelling errors of third grade children's life-letters, Elementary School Journal, Mar. 1938, XXXVIII, 518-27.
- Flesch, Rudolph. Why Johnny can't read. New York: Harper and Row, 1955.
- Fullmer, Daniel W., and Kolson, Clifford J. A beginning reading vocabulary, Journal of Educational Research, Mar. 1961, LIV, 270-2.
- Gates, Arthur I. A reading vocabulary for the primary grades. New York: Teachers College, Columbia University, 1926.
- Gentry, Lillian. A study of the vocabulary load of 66 preprimers, Journal of Educational Research, Mar. 1950, XLIII, 525-32.
- Goldscheider, A., and Muller, R.F. Zur phys. und path. des lesens, Zeitschrift f. klin. med., XXIII, 131. (From Huey, 1908, 435.)
- Greene, Harry A. The new Iowa spelling scale. Iowa City, Iowa: State University of Iowa, 1954.
- Gross, Aline E. Preprimer vocabulary study, Elementary School Journal, Sept. 1934, XXXV, 48-56.
- Harris, Albert J., and Jacobson, Milton D. Basic elementary reading vocabulary. New York: Macmillan Co., 1972.
- Heimberger, Mary J. Analysis and comparison of the vocabularies in textbooks and library books frequently read by early primary children. Unpublished doctoral dissertation, University of Pittsburgh, 1972.
- Hester, Kathleen B. The vocabulary of the sixth grade. Unpublished Ph.D. dissertation, School of Education, University of Pittsburgh, 1938.
- Hitchings, C. Gordon. Comparative difficulty of beginning reading vocabulary: set VI (multisyllabic words). Unpublished doctoral dissertation, University of Pittsburgh, 1977.
- Hockett, John A. Vocabularies of recent primers and first readers, Elementary School Journal, Oct. 1938, XXXIX, 112-15.
- Hollingsworth, Paul M. Spelling lists--outdated? Elementary English, Feb. 1965, XLII, 151-2.
- Horn, Ernest. A basic writing vocabulary: ten thousand words most commonly used in writing. Iowa City, Iowa: State University of Iowa, 1926.
- Horn, Ernest. The commonest words in the spoken vocabulary of children up to and including six years of age, The Twenty-Fourth Yearbook of the National Society for the Study of Education. Bloomington, Illinois: Public School Publishing Company, 1925.

- Horodezky, Betty. Comparative difficulty of beginning reading vocabulary: set II. Unpublished doctoral dissertation, University of Pittsburgh, 1973.
- Huey, Edmund Burke. The psychology and pedagogy of reading. New York: Macmillan, 1908. (Republished--Cambridge, Massachusetts: Massachusetts Institute of Technology, 1968.)
- Johnston, Glenn R. A core vocabulary study in the elementary grades, Elementary English, May 1962, XXXIX, 470-3.
- Jones, W. F. A concrete investigation of the material of English spelling. Vermillion, South Dakota: University of South Dakota, 1914.
- Kolson, Clifford J. The vocabulary of kindergarten children. Unpublished Ph.D. dissertation, School of Education, University of Pittsburgh, 1960.
- Kucera, Henry, and Francis, W. Nelson. Computational analysis of present-day American English. Providence: Brown University Press, 1967.
- Kyte, George C. A core vocabulary for the primary grades, Elementary School Journal, Nov. 1943, XLIV, 157-66.
- Kyte, George C. A core vocabulary in the language arts, Phi Delta Kappan, Mar. 1953, XXXIV, 231-4.
- Losinski, Blanche, and Nemec, Lois G. A study of the difficulty of Dolch basic sight vocabulary in the second and third grades of Wisconsin, Journal of Educational Research, Nov. 1951, XXV, 208-17.
- Mangieri, John N. Difficulty of beginning sight vocabulary taught by three methods. Unpublished doctoral dissertation, University of Pittsburgh, 1972.
- Messmer, Oskar. Zur psychologie des lesens bei kinder und erwachsenen, Archives fur die gesamte psychologie, Dec. 1903, II, 190-298. (From Huey, 1905, 436.)
- Olson, A. V. Analysis of the vocabulary of seven primary reading series, Elementary English, Mar. 1965, XLII, 261-4.
- Pearson, P. David, and Johnson, Dale D. Teaching reading comprehension. New York: Holt, Rinehart and Winston, 1978.
- Penn, Floy. A tentative basic vocabulary for the elementary school. Unpublished Ph.D. dissertation, School of Education, University of Pittsburgh, 1953.
- Reeve, Olive R. The vocabulary of seven primary reading series, Elementary English, Apr. 1958, XXXV, 237-9.

- Rickard, Garrett E. Recognition vocabulary of primary pupils, Journal of Educational Research, Dec. 1935, XXIX, 281-91.
- Rinsland, Henry D. A basic vocabulary of elementary school children. New York: The Macmillan Company, 1947.
- Sartain, Harry W., and Slick, Robert. A comparison of sight vocabulary with the introduction of controlled vocabulary in the basal reading program. Unpublished paper, University of Pittsburgh, Falk Laboratory School, 1968.
- Shambaugh, C. G., and Shambaugh, Olive L. A core vocabulary for elementary school pupils, Journal of Educational Research, Jan. 1929, XIX, 39-46.
- Shambaugh, C. G., and Shambaugh, Olive L. An association study of the vocabulary of grade children, Journal of Educational Research, June 1928, XVIII, 40-47.
- Sholty, Myrtle. A study of reading vocabulary of children. Elementary School Journal, Feb. 1912, XII, 272-97.
- Smith, Madorah E. An investigation of the development of the sentence and the extent of vocabulary in young children. Iowa City, Iowa: University of Iowa, 1926.
- Stone, Clarence R. Second grade reading vocabulary, Elementary School Journal, Jan. 1936, XXXV, 359-67.
- Stone, Clarence R. Graded vocabulary for primary reading. St. Louis: Webster Publishing Co., 1941.
- Story, Suetta B. Does Johnny know more words than Ivan? Reading Teacher, Nov. 1966, 20, 131-133.
- Thorndike, Edward L. The teacher's word book. New York: Teachers College, Columbia University, 1921.
- Thorndike, Edward L. A teacher's word book of the twenty thousand words found most frequently and widely in general reading for children and young people. New York: Teachers College, Columbia University, 1932.
- Thorndike, Edward L., and Lorge, Irving. The teacher's word book of 30,000 words. New York: Teachers College, Columbia University, 1944.
- Trace, Arthur S., Jr. What Ivan knows that Johnny doesn't. New York: Ranjom House, 1961.
- Tidyman, W. F. A survey of the writing vocabularies of public school children in Connecticut. Washington, D.C.: U.S. Bureau of Education, 1921.

- Washburne, Carlton W., and Morphett, Mabel V. Grade placement of children's books, Elementary School Journal, Jan. 1938, XXXVIII, 355-64.
- Wheeler, H. E., and Howell, E. A. First grade vocabulary study, Elementary School Journal, Sept. 1930, XXXI, 52-60.
- Wiley, Will E. Difficult words and the beginner, Journal of Educational Research, Apr. 1928, XVII, 278-86.
- Wilson, Louis A. Children's realistic vocabulary, Elementary English, Jan. 1963, XL, 37-42.
- Zeitler, Julius. Tachistoskopische versuche uber das lesen, Wundt's philosophie studien, SVI (3), 380-463. (From Huey, 1908, 437.)